JOURNAL OF MULTIVARIATE ANALYSIS 12, 306–315 (1982)

An Inequality for the Multivariate Normal Distribution*

LOUIS H. Y. CHEN

National University of Singapore, Singapore, and
Massachusetts Institute of Technology

Communicated by T. L. Lai

Herman Chernoff used Hermite polynomials to prove an inequality for the normal distribution. This inequality is useful in solving a variation of the classical isoperimetric problem which, in turn, is relevant to data compression in the theory of element identification. As the inequality is of interest in itself, we prove a multivariate generalization of it using a different argument.

1. INTRODUCTION

Let X be a random variable having the standard normal distribution and let g be an absolutely continuous real-valued function defined on the real line such that $g(X)$ has finite variance. In [2], Chernoff used Hermite polynomials to prove that

$$\operatorname{Var}[g(X)] \leqslant E[g'(X)]^2 \tag{1.1}$$

and that equality holds if and only if $g(x) = ax + b$ for some constants a and b. As is mentioned in [2], this interesting inequality is useful in solving a variation of the classical isoperimetric problem which, in turn, is relevant to data compression in the theory of element identification (see also Chernoff [1]).

In this paper, we prove a multivariate generalization of inequality (1.1) using a different argument. In our results the finite variance assumption is dropped.

Received October 10, 1980; revised November 1, 1981.
AMS 1980 subject classifications: Primary 60E05, 60E15.
Key words and phrases: Inequality, multivariate normal distribution.
* Research supported in part by National Science Foundation Grant MCS78-01108.

0047-259X/82/020306–10$02.00/0

2. The Main Theorem

We shall adopt the convention that $\int_0^x = -\int_x^0$ if x is negative. The Lebesgue measure in $\mathbb{R}^i$ will be denoted by μ_i for $i \geqslant 1$. We first prove a lemma which is itself a slight extension of Chernoff's inequality in the univariate case.

LEMMA 2.1 *Let X be a random variable having the standard normal distribution. Let ψ and ψ' be functions defined on the real line such that ψ is equal a.e. w.r.t. μ_1 to an indefinite integral of ψ'. Then*

$$\operatorname{Var}[\psi(X)] \leqslant E\left[\int_0^X \psi'(t)\,dt\right]^2 \leqslant E[\psi'(X)]^2. \tag{2.0}$$

Furthermore, $\operatorname{Var}[\psi(X)] = E[\psi'(X)]^2 < \infty$ *if and only if* $\psi(x) = ax + b$ *a.e. w.r.t.* μ_1 *for some constants a and b.*

Proof. Let

$$f(x, t) = I(x > t \geqslant 0) + I(x < t < 0),$$

where $I(A)$ denotes the indicator function of the set A. Then

$$\left(\int_0^x \psi'(t)\,dt\right)^2 = \left(\int_{-\infty}^{\infty} f(x, t)\,\psi'(t)\,dt\right)^2$$

and by the Cauchy–Schwarz inequality

$$\begin{aligned}\left(\int_0^x \psi'(t)\,dt\right)^2 &\leqslant \left(\int_{-\infty}^{\infty} f(x, t)\,dt\right)\left(\int_{-\infty}^{\infty} f(x, t)[\psi'(t)]^2\,dt\right)\\ &= |x| \int_{-\infty}^{\infty} f(x, t)[\psi'(t)]^2\,dt\end{aligned}$$

where it is noted that $f^2 = f$. Therefore,

$$\begin{aligned}E\left(\int_0^X \psi'(t)\,dt\right)^2 &\leqslant E\,|X| \int_{-\infty}^{\infty} f(X, t)[\psi'(t)]^2\,dt\\ &= \int_{-\infty}^{\infty} [E\,|X|\,f(X, t)][\psi'(t)]^2\,dt\\ &= \int_{-\infty}^{\infty} (2\pi)^{-1/2}\,e^{-t^2/2}[\psi'(t)]^2\,dt\\ &= E[\psi'(X)]^2.\end{aligned}$$

This proves the second inequality in (2.0). The first inequality in (2.0) follows from the assumption that $\psi(x)$ differs from $\int_0^x \psi'(t)\,dt$ by a constant for almost all x w.r.t. μ_1.

If $\psi(x) = ax + b$ a.e. w.r.t. μ_1 for some constants a and b, then trivially $\operatorname{Var}[\psi(X)] = E[\psi'(X)]^2 < \infty$. Conversely, suppose the latter holds. Then using the fact that the standard normal distribution is equivalent to μ_1, we have

$$\left(\int_{-\infty}^{\infty} f(x,t)\,\psi'(t)\,dt\right)^2 = \left(\int_{-\infty}^{\infty} f(x,t)\,dt\right)\left(\int_{-\infty}^{\infty} f(x,t)[\psi'(t)]^2\,dt\right)$$

for almost all x w.r.t. μ_1. This implies that for almost all x w.r.t. μ_1 $f(x,t)\,\psi'(t) = a(x)\,f(x,t)$ or $f(x,t) = 0$ for almost all t w.r.t. μ_1, where $a(x)$ depends only on x. It is not difficult to see that the second alternative that $f(x,t) = 0$ for almost all t w.r.t. μ_1 is not possible for x belonging to a set of positive measure. So for almost all x w.r.t. μ_1 we have $\psi'(t) = a(x)$ for almost all t w.r.t. μ_1 such that $0 < t \leqslant x$ or $t < x < 0$. This implies that for almost all t w.r.t. μ_1,

$$\begin{aligned} \psi'(t) &= a_1 \qquad \text{if} \quad t \geqslant 0 \\ &= a_2 \qquad \text{if} \quad t < 0, \end{aligned}$$

where a_1 and a_2 are constants. Therefore, $\psi = \psi^{\sim}$ a.e. w.r.t. μ_1 where

$$\begin{aligned} \psi^{\sim}(x) &= a_1 x + b \qquad \text{if} \quad x \geqslant 0 \\ &= a_2 x + b \qquad \text{if} \quad x < 0, \end{aligned}$$

and b is a constant. But the equality $\operatorname{Var}[\psi(X)] = E[\psi'(X)]^2 < \infty$ also implies that

$$\begin{aligned} \operatorname{Var}[\psi^{\sim}(X)] = \operatorname{Var}[\psi(X)] &= E\left[\int_0^X \psi'(t)\,dt\right]^2 \\ &= E[\psi^{\sim}(X) - \psi^{\sim}(0)]^2. \end{aligned}$$

Therefore, $E\psi^{\sim}(X) = \psi^{\sim}(0) = b$. Consequently, $a_1 = a_2$, and we have $\psi(x) = ax + b$ a.e. w.r.t. μ_1 for some constants a and b. This completes the proof of the lemma.

THEOREM 2.1. *Let $X_1, \ldots, X_k$ be independent random variables each having the standard normal distribution. Let $g, g_1, \ldots, g_k$ be real-valued Borel measurable functions defined on $\mathbb{R}^k$ such that*

$$\begin{aligned} g(x_1, \ldots, x_k) = &\int_0^{x_i} g_i(x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_k)\,dt \\ &+ g(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_k) \end{aligned} \tag{2.1}$$

a.e. w.r.t. μ_k *for* $i=1,\dots,k$. *Then*

$$\operatorname{Var}[g(X_1,\dots,X_k)] \leqslant \sum_{i=1}^{k} E[g_i(X_1,\dots,X_k)]^2, \tag{2.2}$$

where equality holds with the right-hand side finite if and only if $g(x_1,\dots,x_k)=a_1x_1+\cdots+a_kx_k+b$ *a.e. w.r.t.* μ_k *for some constants* $a_1,\dots,a_k$ *and* b.

Proof. If $k=1$, the theorem follows trivially from Lemma 2.1. So let $k\geqslant 2$. There is nothing to prove if $\sum_{i=1}^{k} E[g_i(X_1,\dots,X_k)]^2=\infty$. So we assume it to be finite. The first thing to do is to show that

$$E[g(X_1,\dots,X_k)]^2<\infty. \tag{2.3}$$

We do this by induction. By Fubini's theorem,

$$E[g_i(x_1,\dots,x_{i-1},X_i,\dots,X_k)]^2<\infty \tag{2.4}$$

for almost all $(x_1,\dots,x_{i-1})$ w.r.t. μ_{i-1} and for $i=1,\dots,k$. Here and throughout the rest of the proof, the statement "for almost all $(x_1,\dots,x_{i-1})$ w.r.t. μ_{i-1}" is to be omitted for $i=1$. By Lemma 2.1 and the independence of $X_1,\dots,X_k$ we have

$$E\left[\int_0^{X_i} g_i(x_1,\dots,x_{i-1},t,X_{i+1},\dots,X_k)\,dt\right]^2$$
$$\leqslant E[g_i(x_1,\dots,x_{i-1},X_i,\dots,X_k)]^2 \tag{2.5}$$

for almost all $(x_1,\dots,x_{i-1})$ w.r.t. μ_{i-1} and for $i=1,\dots,k$. We start with $i=k$. From condition (2.1), we have, for almost all $(x_1,\dots,x_{k-1})$ w.r.t. μ_{k-1},

$$g(x_1,\dots,x_{k-1},X_k)=g(x_1,\dots,x_{k-1},0)$$
$$+\int_0^{X_k} g_k(x_1,\dots,x_{k-1},t)\,dt \qquad \text{a.s.}$$

Combining this with (2.5) we obtain

$$E[g(x_1,\dots,x_{k-1},X_k)]^2<\infty$$

for almost all $(x_1,\dots,x_{k-1})$ w.r.t. μ_{k-1}. Now assume that

$$E[g(x_1,\dots,x_i,X_{i+1},\dots,X_k)]^2<\infty \tag{2.6}$$

for almost all $(x_1, ..., x_i)$ w.r.t. μ_i. Since the univariate normal density function is bounded away from zero on any bounded interval, it follows from (2.4) that

$$E\left|\int_0^{x_i} [g_i(x_1, ..., x_{i-1}, t, X_{i+1}, ..., X_k)]^2 \, dt\right| < \infty \tag{2.7}$$

for almost all $(x_1, ..., x_{i-1})$ w.r.t. μ_{i-1} and all x_i. This together with (2.1) and (2.6) imply that

$$E[g(x_1, ..., x_{i-1}, 0, X_{i+1}, ..., X_k)]^2 < \infty \tag{2.8}$$

for almost all $(x_1, ..., x_{i-1})$ w.r.t. μ_{i-1}. Combining (2.1), (2.5) and (2.8), we obtain

$$E[g(x_1, ..., x_{i-1}, X_i, ..., X_k)]^2 < \infty$$

for almost all $(x_1, ..., x_{i-1})$ w.r.t. μ_{i-1}. This proves (2.3).

Now let $\mathscr{F}_0$ be the trivial σ-algebra and $\mathscr{F}_i$ be the σ-algebra generated by $X_1, ..., X_i$, $i = 1, ..., k$. Write E_i for the conditional expectation operator given $\mathscr{F}_i$, $i = 0, 1, ..., k$. Define $Y_i = E_i\, g(X_1, ..., X_k) - E_{i-1}\, g(X_1, ..., X_k)$ for $i = 1, ..., k$. Then $\{Y_i : i = 1, ..., k\}$ is a sequence of martingale differences relative to $\{\mathscr{F}_i : i = 1, ..., k\}$ and $g(X_1, ..., X_k) - Eg(X_1, ..., X_k) = \sum_{i=1}^k Y_i$. By the martingale property,

$$\mathrm{Var}[g(X_1, ..., X_k)] = \sum_{i=1}^k EY_i^2. \tag{2.9}$$

But by the independence of $X_1, ..., X_k$, Eq. (2.1) and Lemma 2.1, we have for $i = 1, ..., k$,

$$\begin{aligned} EY_i^2 &= E\left[E_i \int_0^{X_i} g_i(X_1, ..., X_{i-1}, t, X_{i+1}, ..., X_k)\, dt \right. \\ &\qquad \left. - E_{i-1} \int_0^{X_i} g_i(X_1, ..., X_{i-1}, t, X_{i+1}, ..., X_k)\, dt\right]^2 \\ &\leqslant E\left[E_i \int_0^{X_i} g_i(X_1, ..., X_{i-1}, t, X_{i+1}, ..., X_k)\, dt\right]^2 \\ &= EE_{i-1}\left[\int_0^{X_i} E_i\, g_i(X_1, ..., X_{i-1}, t, X_{i+1}, ..., X_k)\, dt\right]^2 \\ &\leqslant EE_{i-1}[E_i\, g_i(X_1, ..., X_k)]^2 \\ &\leqslant E[g_i(X_1, ..., X_k)]^2. \end{aligned} \tag{2.10}$$

This proves (2.2).

If $g(x_1,\ldots,x_k)=a_1x_1+\cdots+a_kx_k+b$ a.e. w.r.t. μ_k, then inequality (2.2) trivially reduces to an equality with both sides finite. Conversely, if equality in (2.2) holds with the right-hand side finite, then by (2.9) and (2.10), $EY_i^2 = E[g_i(X_1,\ldots,X_k)]^2<\infty$ for $i=1,\ldots,k$, which, in turn, implies that

$$\begin{aligned}\mathrm{Var}_{i-1}&\left[\int_0^{X_i} E_i\, g_i(X_1,\ldots,X_{i-1},t,X_{i+1},\ldots,X_k)\,dt\right]^2\\ &= E_{i-1}[E_i\, g_i(X_1,\ldots,X_k)]^2 \qquad \text{a.s.,}\end{aligned}$$

where Var_{i-1} denotes conditional variance given $\mathscr{F}_{i-1}$. By Lemma 2.1,

$$\int_0^{X_i} E_i\, g_i(X_1,\ldots,X_{i-1},t,X_{i+1},\ldots,X_k)\,dt = u_iX_i+v_i \qquad \text{a.s.,}$$

where u_i and v_i are $\mathscr{F}_{i-1}$ measurable, $i=1,\ldots,k$. It follows that for $i=1,\ldots,k$,

$$Y_i = u_iX_i \qquad \text{a.s.}$$

We now prove by induction that all the u_i must be constants a.s. We have proved that u_1 is a constant a.s. Suppose $u_1,\ldots,u_j$ are constants a.s., say $u_1=a_1,\ldots,u_j=a_j$ a.s., $1\leqslant j<k$. Then

$$\begin{aligned}g&(X_1,\ldots,X_k)-Eg(X_1,\ldots,X_k)\\ &= a_1X_1+\cdots+a_jX_j+u_{j+1}X_{j+1}+\cdots+u_kX_k \qquad \text{a.s.} \qquad (2.11)\end{aligned}$$

Now rearrange $X_1,\ldots,X_k$ in the order X_{j+1}, $X_1,\ldots,X_j$, $X_{j+2},\ldots,X_k$. Define $\mathscr{F}_i^{\sim}$, $E_i^{\sim}$, $Y_i^{\sim}$ with respect to $X_{j+1}, X_1,\ldots,X_j, X_{j+2},\ldots,X_k$ in the same way as $\mathscr{F}_i$, E_i, Y_i are with respect to $X_1,\ldots,X_k$, $i=1,\ldots,k$. For example, $\mathscr{F}_1^{\sim}$, $\mathscr{F}_2^{\sim}$, etc., are respectively σ-algebras generated by $\{X_{j+1}\}$, $\{X_{j+1},X_1\}$, etc. Then

$$\begin{aligned}Y_1^{\sim} &= E_1^{\sim}(a_1X_1+\cdots+a_jX_j+u_{j+1}X_{j+1}+\cdots+u_kX_k)\\ &= E(a_1X_1+\cdots+a_jX_j+u_{j+1}X_{j+1}+\cdots+u_kX_k \mid X_{j+1})\\ &= (Eu_{j+1})\,X_{j+1} \qquad \text{a.s.}\end{aligned}$$

and

$$E(Y_1^{\sim})^2 = (Eu_{j+1})^2.$$

But by arguments similar to those leading to (2.9) and (2.10), the equality in (2.2) implies that

$$\begin{aligned}E(Y_1^{\sim})^2 &= E[g_{j+1}(X_1,\ldots,X_k)]^2\\ &= EY_{j+1}^2 = Eu_{j+1}^2.\end{aligned}$$

It follows that

$$Eu_{j+1}^2 = (Eu_{j+1})^2$$

which is $\mathrm{Var}(u_{j+1}) = 0$ and hence implies that u_{j+1} is a constant a.s. This proves that all the u_i are constants a.s. Consequently,

$$g(x_1,..., x_k) = a_1 x_1 + \cdots + a_k x_k + b \qquad \text{a.e. w.r.t. } \mu_k,$$

where $a_1,..., a_k$ and b $(=Eg(X_1,..., X_k))$ are constants. This completes the proof of the theorem.

3. COROLLARIES

In this section we consider inequality (2.2) for two specific classes of functions which satisfy (2.1).

COROLLARY 3.1. *Let $X_1,..., X_k$ be independent random variables each having the standard normal distribution. Let g be a partially differentiable real-valued function defined on $\mathbb{R}^k$. Then*

$$\mathrm{Var}[g(X_1,..., X_k)] \leqslant \sum_{i=1}^{k} E[D_i\, g(X_1,..., X_k)]^2,$$

where equality holds with the right-hand side finite if and only if

$$g(x_1,..., x_k) = a_1 x_1 + \cdots + a_k x_k + b$$

for all $x_1,..., x_k$, and some constants $a_1,..., a_k$ and b.

We note that partially differentiable functions are separately continuous (that is, continuous in each variable) and that separately continuous functions from $\mathbb{R}^k$ to $\mathbb{R}$ are Borel measurable (see, for example, Kuratowski [3, p. 285]). The proof of Corollary 3.1 follows trivially from these facts and the following lemma.

LEMMA 3.1. *Let $h: \mathbb{R}^k \to \mathbb{R}$ be separately continuous. If $h = 0$ a.e. w.r.t. μ_k, then $h = 0$ everywhere.*

Proof. Suppose there is a point at which $h \neq 0$. Without loss of generality, we may assume that $h(0,..., 0) = 2$. By the separate continuity of h, there exist an open interval J_1 containing 0 such that

$$h(x_1, 0,..., 0) > 1 \qquad \text{for} \quad x_1 \in J_1.$$

For each $x_1 \in J_1$, there exists an open interval $J_2(x_1)$ containing 0 such that

$$h(x_1, x_2, 0,..., 0) > \tfrac{1}{2} \qquad \text{for} \quad x_2 \in J_2(x_1).$$

Arguing inductively, we have

$$h(x_1,..., x_k) > \frac{1}{k}\cdot$$

for $x_i \in J_i(x_1,..., x_{i-1})$ for $i = 1,..., k$ where each $J_i(x_1,..., x_{i-1})$ is an open interval containing 0 and $J_i(x_1,..., x_{i-1})$ is defined to be J_1 for $i = 1$. Let $A = \{(x_1,..., x_k): x_i \in J_i(x_1,..., x_{i-1}),\quad i = 1,..., k\}$. Clearly $A \subset \{(x_1,..., x_k): h(x_1,..., x_k) > 1/k\}$. But $h = 0$ a.e. w.r.t. μ_k. This implies that $\mu_k(h > 1/k) = 0$ which, in turn, implies that A is Lebesgue measurable and $\mu_k(A) = 0$. Now define

$$A(n_2,..., n_k) = \left\{(x_1,..., x_k) \in A: \mu_1(J_j(x_1,..., x_{j-1})) > \frac{1}{n_j}, j = 2,..., k\right\}.$$

Then $A = \bigcup_{n_2=1}^{\infty} \cdots \bigcup_{n_k=1}^{\infty} A(n_2,..., n_k)$ so that for each $(n_2,..., n_k)$, $A(n_2,..., n_k)$ is Lebesgue measurable and $\mu_k(A(n_2,..., n_k)) = 0$. Define

$$B(n_2,..., n_k) = \{x_1 : (x_1,..., x_k) \in A(n_2,..., n_k)\}.$$

Clearly $J_1 = \bigcup_{n_2=1}^{\infty} \cdots \bigcup_{n_k=1}^{\infty} B(n_2,..., n_k)$ and $B(n_2,..., n_k)$ is nondecreasing in $n_2,..., n_k$. Therefore, for sufficiently large $n_2,..., n_k$, $\mu_1(B(n_2,..., n_k)) \geqslant \frac{1}{2}\mu_1(J_1) > 0$. But for sufficiently large $n_2,..., n_k$, we have by Fubini's theorem

$$\begin{aligned} 0 = \mu_k(A(n_2,..., n_k)) &= \underset{(x_1,...,x_k)\in A(n_2,...,n_k)}{\int \cdots \int} d\mu_1(x_k) \cdots d\mu_1(x_1) \\ &\geqslant \frac{1}{n_2 \cdots n_k} \int_{x_1 \in B(n_2,...,n_k)} d\mu_1(x_1) \\ &\geqslant \frac{\mu_1(J_1)}{2n_2 \cdots n_k} > 0 \end{aligned}$$

which is a contradiction. This proves the lemma.

If g has a differential at every point in $\mathbb{R}^k$, an inequality for the multivariate normal distribution with zero mean vector and any covariance matrix Σ can be deduced from Corollary 3.1. For the next corollary, let m $(1 \leqslant m \leqslant k)$ be the rank of Σ and $\mathscr{L}(\Sigma)$ denote the real linear space spanned by the column vectors of Σ. Let $v_1,..., v_m$ be column vectors which form a basis of $\mathscr{L}(\Sigma)$ and g_Σ be the restriction of g to $\mathscr{L}(\Sigma)$. The gradient of g will be denoted by ∇g and be represented by a column vector. The transpose of a matrix A will be denoted by A^t.

COROLLARY 3.2. *Let $\xi = (X_1, ..., X_k)^t$ be a random vector having the multivariate normal distribution with zero mean vector and covariance matrix Σ. Let g be a real-valued function defined on $\mathbb{R}^k$ such that it has a differential at every point. Then*

$$\operatorname{Var}[g(\xi)] \leqslant E\nabla^t g(\xi)\, \Sigma \nabla g(\xi), \tag{3.1}$$

where equality holds with the right-hand side finite if and only if $g_\Sigma(x_1 v_1 + \cdots + x_m v_m) = a_1 x_1 + \cdots + a_m x_m + b$ for all real numbers $x_1, ..., x_m$ and some constants $a_1, ..., a_m$ and b.

Proof. We note that the form of g_Σ corresponding to equality in (3.1) is independent of the choice of the basis $\{v_1, ..., v_m\}$ of $\mathscr{L}(\Sigma)$. Thus, we shall choose $v_1, ..., v_m$ to be such that $\xi = Y_1 v_1 + \cdots + Y_m v_m$ with $Y_1, ..., Y_m$ being independent random variables each having the standard normal distribution. Let $v_{m+1}, ..., v_k$ be linearly independent column vectors such that $\{v_1, ..., v_k\}$ is a basis of $\mathbb{R}^k$. Let $C = (v_1, ..., v_m)$ and $A = (v_1, ..., v_k) = (c_{ij})$. Then $\Sigma = CC'$. Consider the mapping $f: \mathbb{R}^k \to \mathbb{R}^k$ defined by $f(y) = Ay$, $y \in \mathbb{R}^k$. Define $h: \mathbb{R}^k \to \mathbb{R}$ by $h = g \circ f$. Since g has a differential, so has h and

$$\nabla h(y) = \sum_{i=1}^{k} D_i g(f(y))(c_{i1}, ..., c_{ik})^t.$$

Therefore,

$$\begin{aligned}
\nabla^t g(\xi)\, \Sigma \nabla g(\xi) &= \nabla^t g(\xi)\, CC' \nabla g(\xi) \\
&= \sum_{j=1}^{m} [v_j^t \nabla g(\xi)]^2 = \sum_{j=1}^{m} \left[\sum_{i=1}^{k} c_{ij} D_i g(\xi) \right]^2 \\
&= \sum_{j=1}^{m} [D_j h(Y_1, ..., Y_m, 0, ..., 0)]^2.
\end{aligned} \tag{3.2}$$

There is nothing to prove if $E\nabla^t g(\xi)\, \Sigma \nabla g(\xi) = \infty$. However, if $E\nabla^t g(\xi)\, \Sigma \nabla g(\xi) < \infty$, then, by (3.2), $\sum_{j=1}^{m} E[D_j h(Y_1, ..., Y_m, 0, ..., 0)]^2 < \infty$. Applying Corollary 3.1, we obtain

$$\begin{aligned}
\operatorname{Var}[g(\xi)] &= \operatorname{Var}[h(Y_1, ..., Y_m, 0, ..., 0)] \\
&\leqslant \sum_{j=1}^{m} E[D_j h(Y_1, ..., Y_m, 0, ..., 0)]^2 \\
&= E\nabla^t g(\xi)\, \Sigma \nabla g(\xi).
\end{aligned}$$

It is easy to check that if $g_\Sigma(x_1 v_1 + \cdots + x_m v_m) = a_1 x_1 + \cdots + a_m x_m + b$ for some constants $a_1,..., a_m$ and b, then (3.1) reduces to an equality with the right-hand side finite. Conversely, if the latter holds, then

$$\mathrm{Var}[h(Y_1,..., Y_m, 0,..., 0)] = \sum_{j=1}^{m} E[D_j h(Y_1,..., Y_m, 0,..., 0)]^2$$

and by Corollary 3.1 again

$$h(x_1,..., x_m, 0,..., 0) = a_1 x_1 + \cdots + a_m x_m + b$$

for some constants $a_1,..., a_m$ and b. But

$$h(x_1,..., x_m, 0,..., 0) = g_\Sigma(x_1 v_1 + \cdots + x_m v_m).$$

This completes the proof of the corollary.

Acknowledgments

The author is thankful to Dr. Huei Chuen Huang for a discussion which led to the proof of Lemma 3.1. He is also thankful to the referee for his suggestions which helped to simplify the proofs of Lemma 2.1 and Theorem 2.1.

References

[1] Chernoff, H. (1980). The identification of an element of a large population in the presence of noise. *Ann. Statist.* **8** 1179–1197.

[2] Chernoff, H. (1981). A note on an inequality involving the normal distribution. *Ann. Probab.* **9** 533–535.

[3] Kuratowski, C. (1958). *Topologie.* Panstwowe Wydawnictwo Naukowe, Warszawa.